#5 Time Series & Data Visualization

by Saturdays.AI

Saturdays.AI

Machine Learning

Saturdays.AI
Donostia

Schedule

State of the course

Session 5 Review

Challenge

Notebook + resources

State of the course

Saturdays.AI
Donostia

#1 Cleaning & Exploratory Data Analysis ✔️

#2 Supervised Learning ✔️

#3 Decision Trees & Random Forest ✔️

#4 Unsupervised Learning: Clustering & Dim. Red. ✔️

#5 Time Series Analysis + Data Viz 🔴 Today!

#6 Neural Networks, Gradient Descent 🔜

#7 NLP 🔜

Time Series & Data Visualization

Saturdays.AI
Donostia

Time series

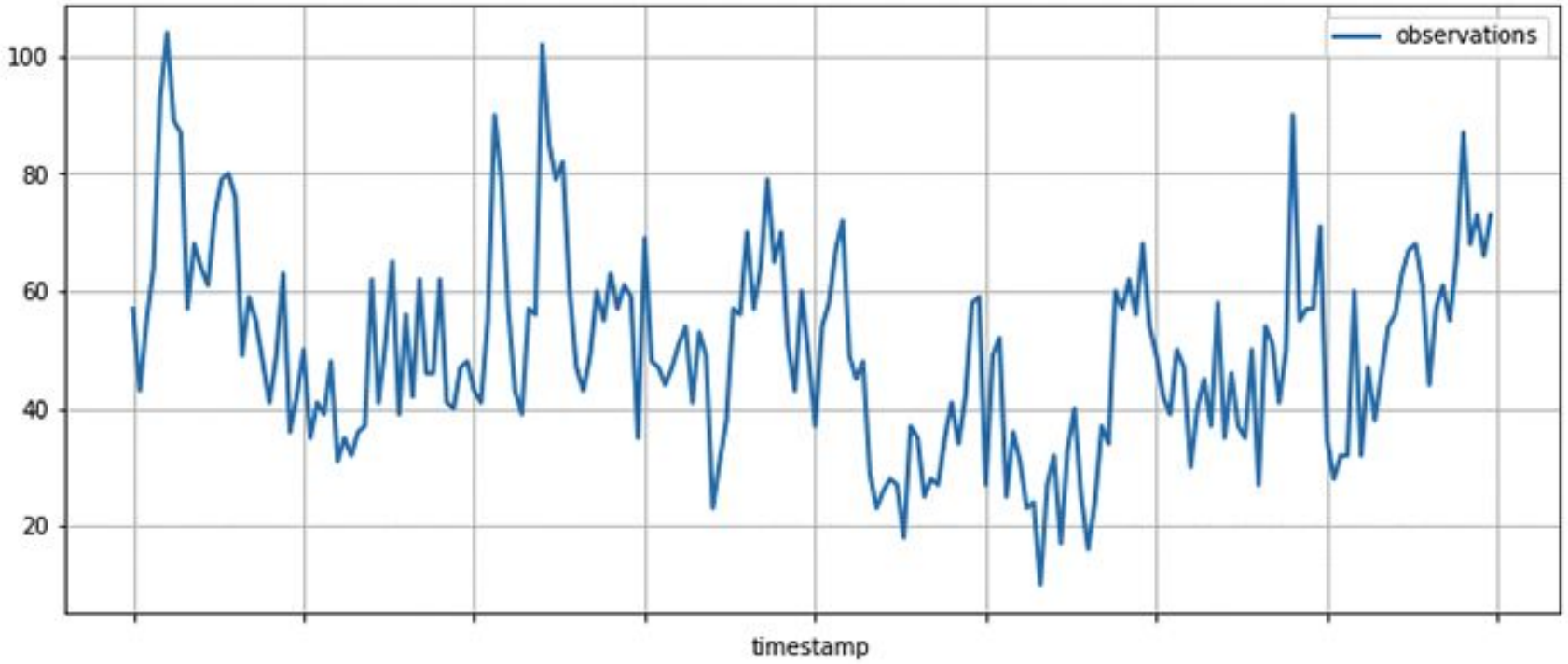
- Definitions
- Modeling
    - Components
    - Autocorrelation
    - ARIMA
    - Prediction Intervals
    - B - Plans

Data Visualization

- Small Review
- Tips
- DO NOT's

DEFINITIONS

Saturdays.AI
Donostia



What is a time series

Saturdays.AI
Donostia

- **Data points listed in a time order**
- **A sequence is taken at successive equally spaced points**
- **Is a continuous variable**

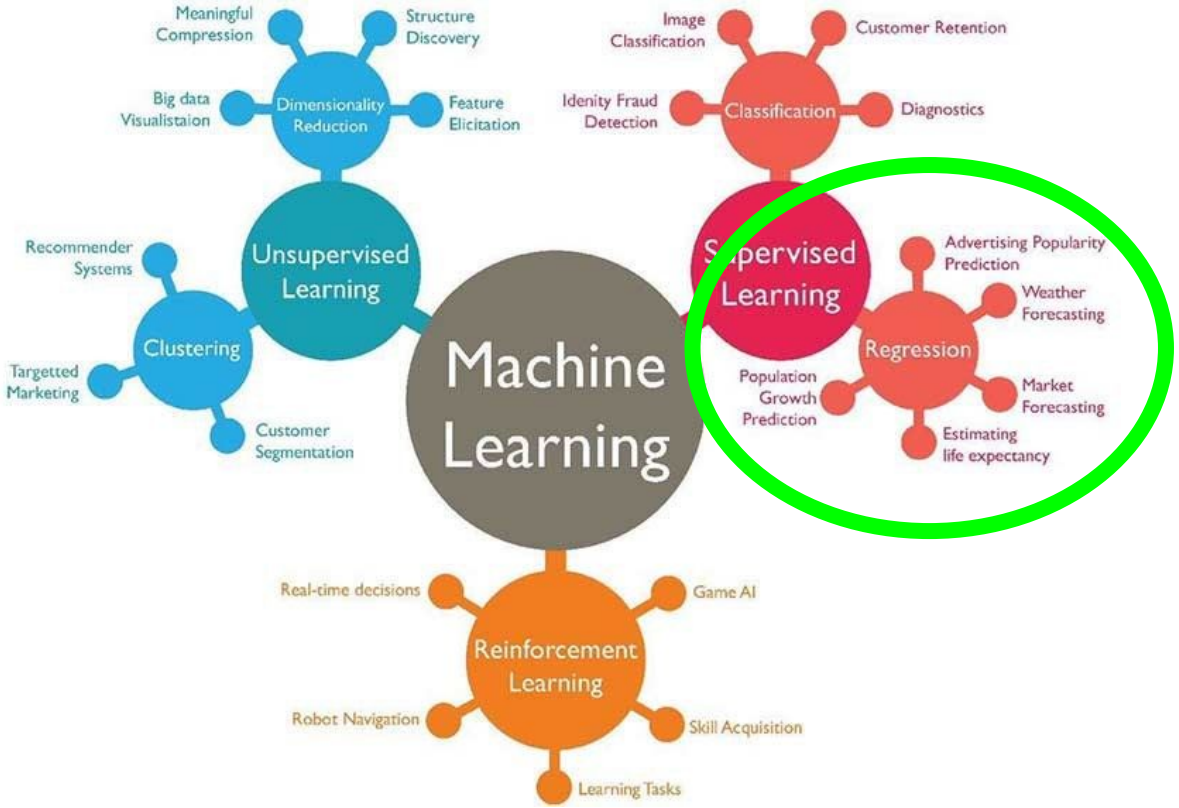
  - Temperatures, tides, ratios, arrivals …

What is NOT time series

Saturdays.AI
Donostia

- **Anything that has a timestamp**
  - Some variables can be treated like a time series if EDA is performed, but not all of them
  - Categorical variables: heights
  - Not evenly spaced variables: Counting how many people enter a bar

What should NOT be modeled as a time series

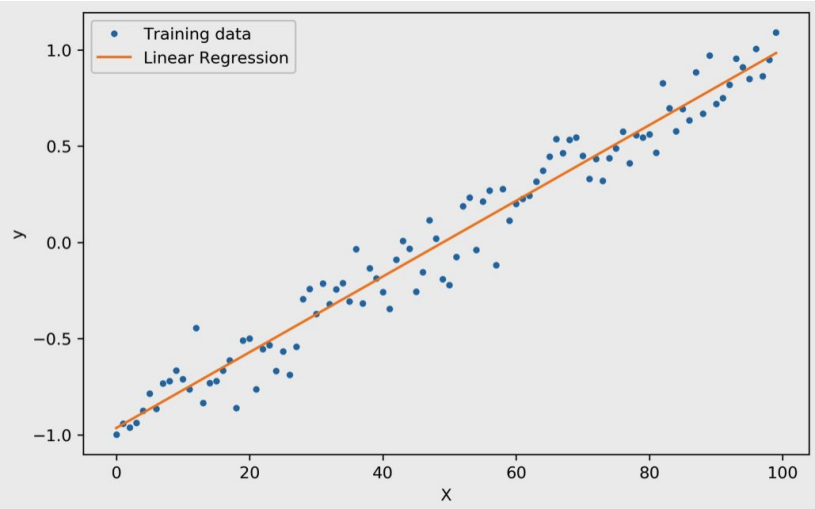


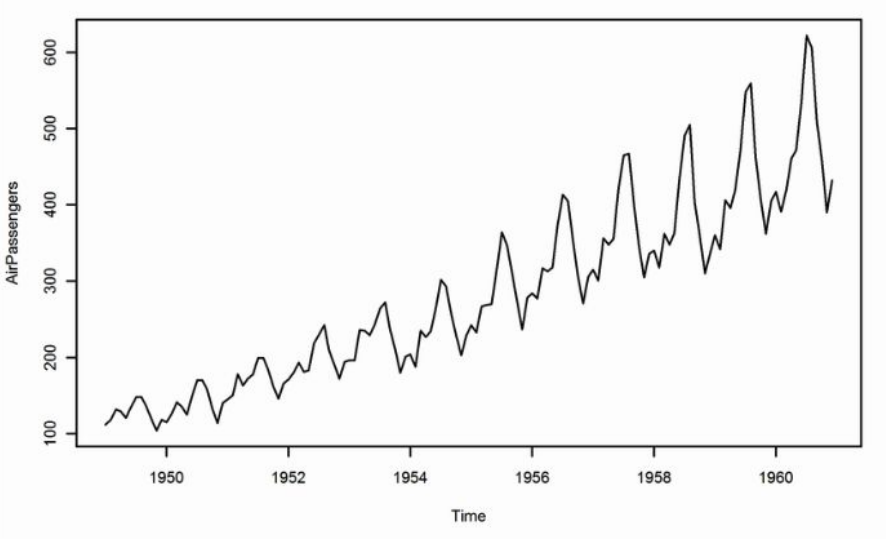
Saturdays.AI
Donostia

What is the difference between a Regression and a Time Series

Saturdays.AI
Donostia

TIME = ORDER

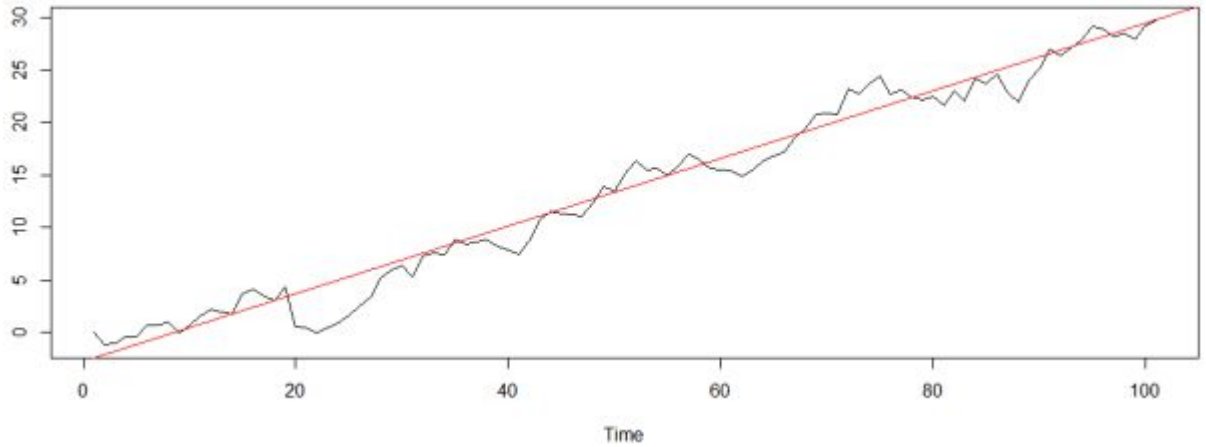




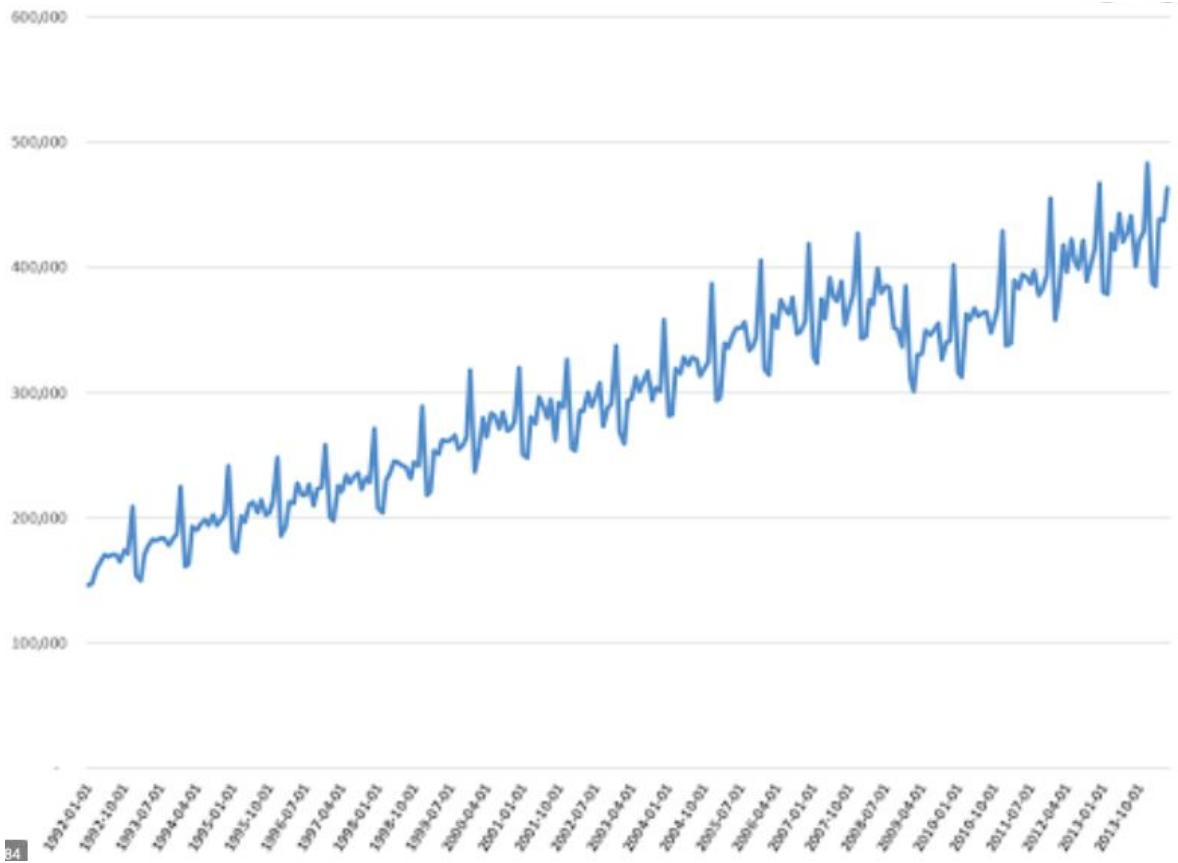
MODELING

Research question
Data validation + cleaning
Exploratory analysis
Modeling phase
Performance

Components: Trend



Components: Seasonality

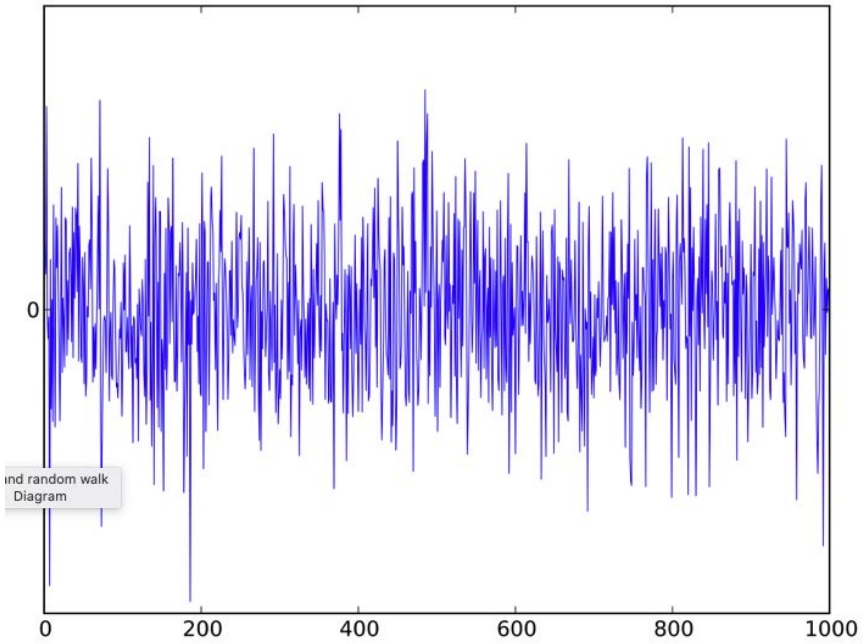


Saturdays.AI
Donostia

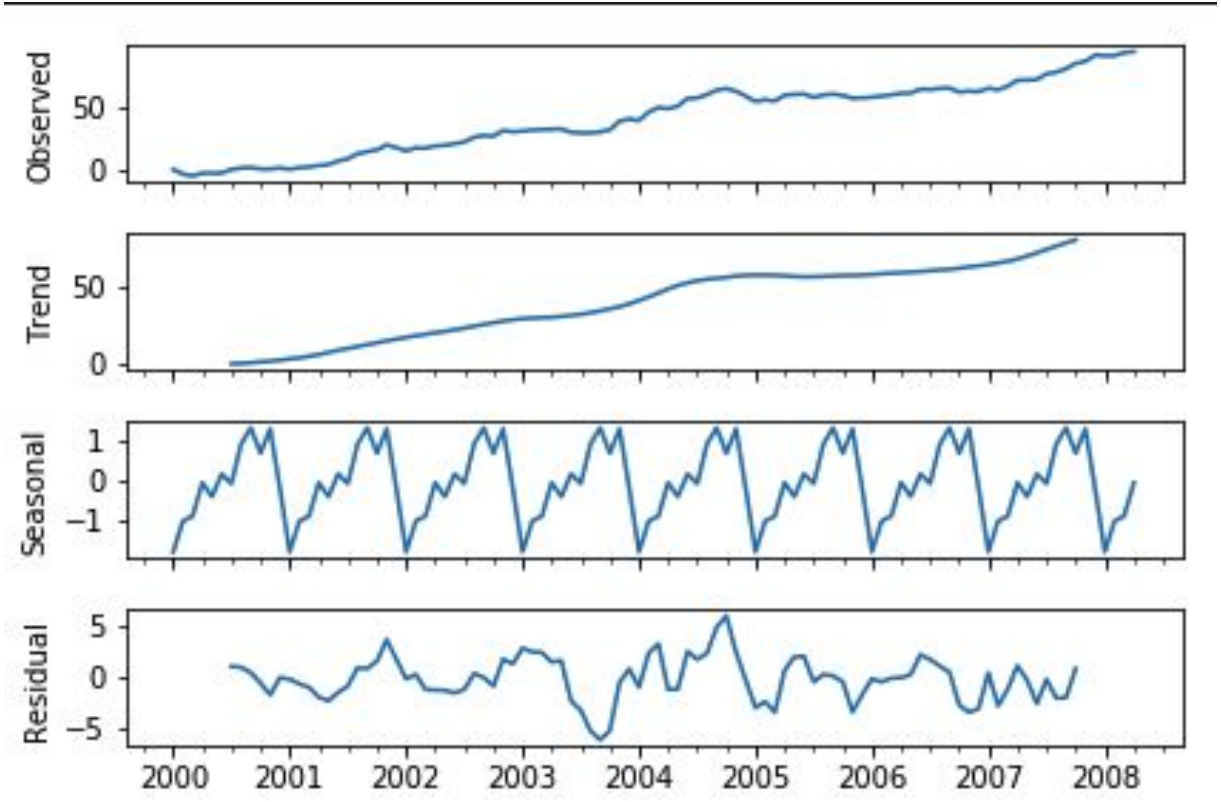
Components: Random

Saturdays.AI
Donostia

- **White noise**
  - **Random white noise**



Components: Decomposition

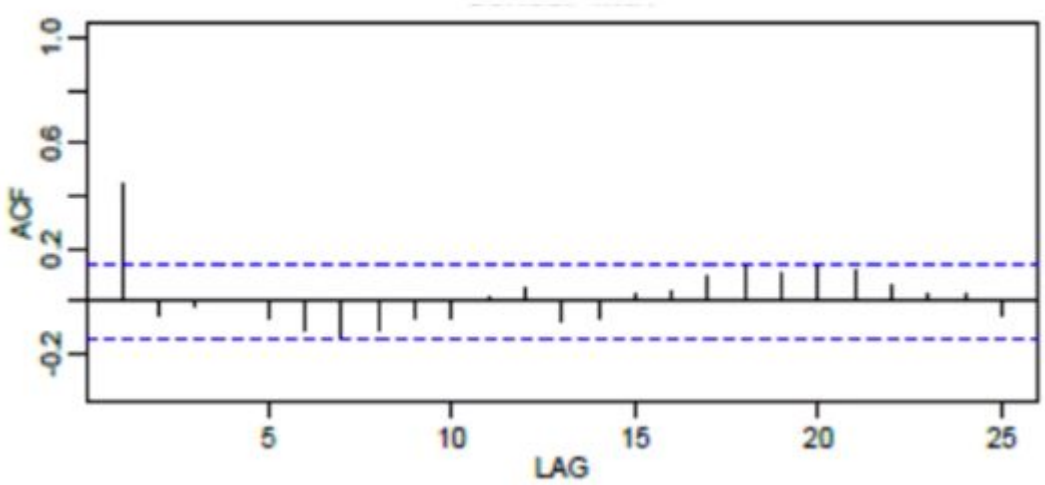


Saturdays.AI
Donostia

Autocorrelation: Autocorrelation

Saturdays.AI
Donostia

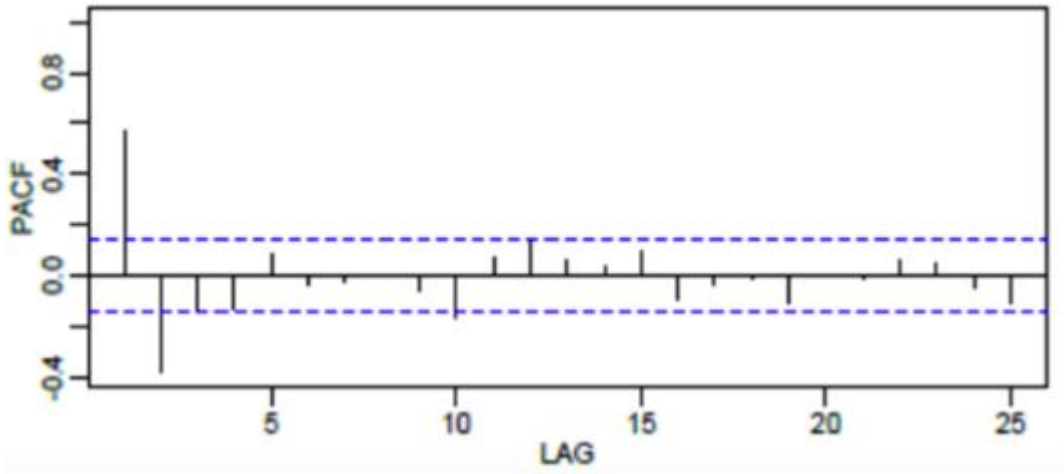
- **The correlation (linear dependence) between one observation and all the previous points in time.**



Autocorrelation: Partial Autocorrelation

Saturdays.AI
Donostia

- Correlation between two observations made at different times



Autocorrelation: Stationarity (I)

Saturdays.AI
Donostia

Basic assumptions:

- The time series has certain regularity along time
- The time series is stationary (or at least we can transform it to one)

Autocorrelation: Stationarity (II)

Saturdays.AI
Donostia

What does it mean to be stationary:

- The mean value of the time series is NOT a function of time.
- The variance is NOT a function of time

- The covariance of a point i and a point i+t, is NOT a function of time

Autocorrelation: Stationarity (III)

Saturdays.AI
Donostia

How do we know whether a time series is stationary or not:

- Dickey-Fuller test
  - p-value < 0.05
  - ADF & 1%, 5%, 10%, as close as possible

NO STATIONARY

```
ADF Statistic: 0.8153688792060423
p-value: 0.9918802434376409
Critical Values:
        1%: -3.4816817173418295
        5%: -2.8840418343195267
        10%: -2.578770059171598
```

ARIMA: AR (p)

Saturdays.AI
Donostia

- Premise: the past values have an effect on the current values
- The order (p), corresponds to the number of days incorporated on the formula.
- We can get (p) from the partial autocorrelations

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \cdots + \beta_p y_{t-p}$$

ARIMA: MA(q)

Saturdays.AI
Donostia

- Value of the current day, depends on the errors of the past days
- q is the number of terms to be included
- The Autocorrelation function helps to get the best (q) parameter

$$y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q}$$

ARIMA: ARMA(p, q)

Saturdays.AI
Donostia

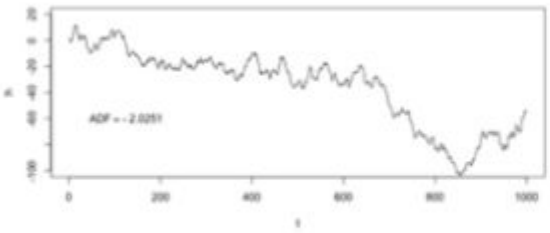
- Adds the two models seen before = AR + MA

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \cdots + \beta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q}$$

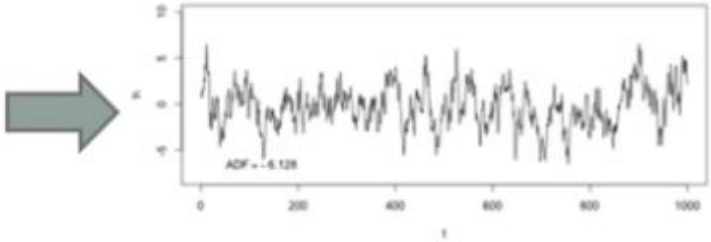
ARIMA: (p, d, q)

Saturdays.AI
Donostia

- Adds a (d) differencing component to ARMA
- Differencing subtracts the current value from the previous.
- It can be used to transform a time series into one that is stationary



Data with trend

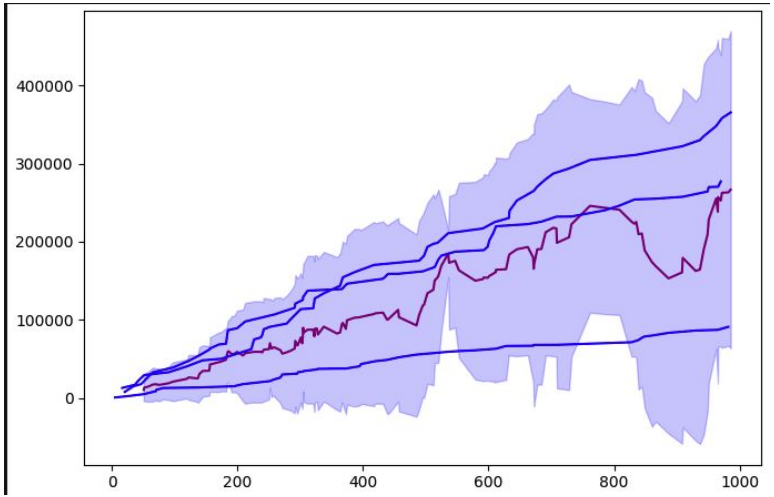


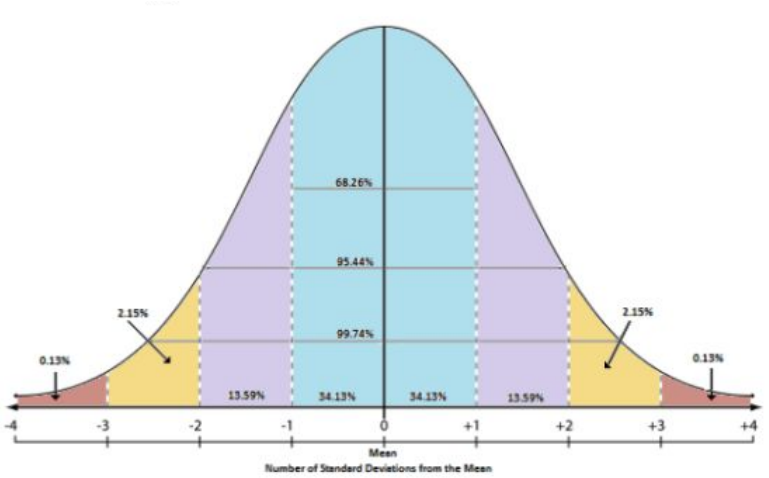
Data after one-lag differencing

Prediction Intervals

Saturdays.AI Donostia

Rolling means and standard deviation



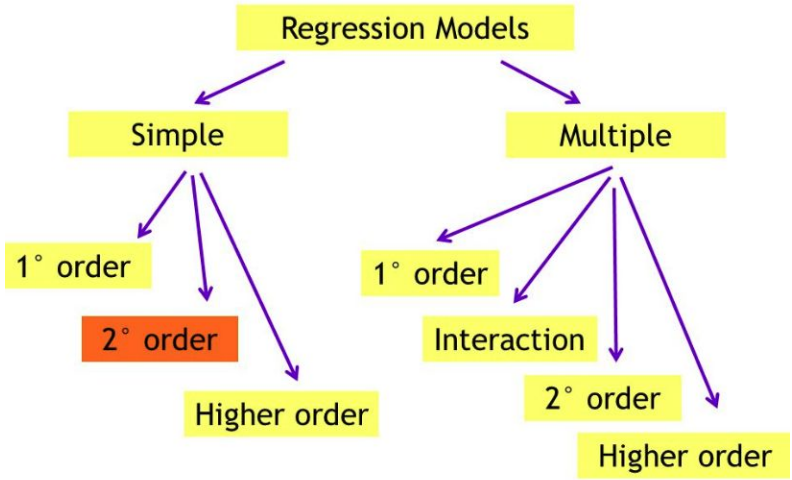


Source: deltanalytics.org

B-Plan : Diagram

1. Is our data a Time series?
2. Is it stationary? Can we transform it to stationary?

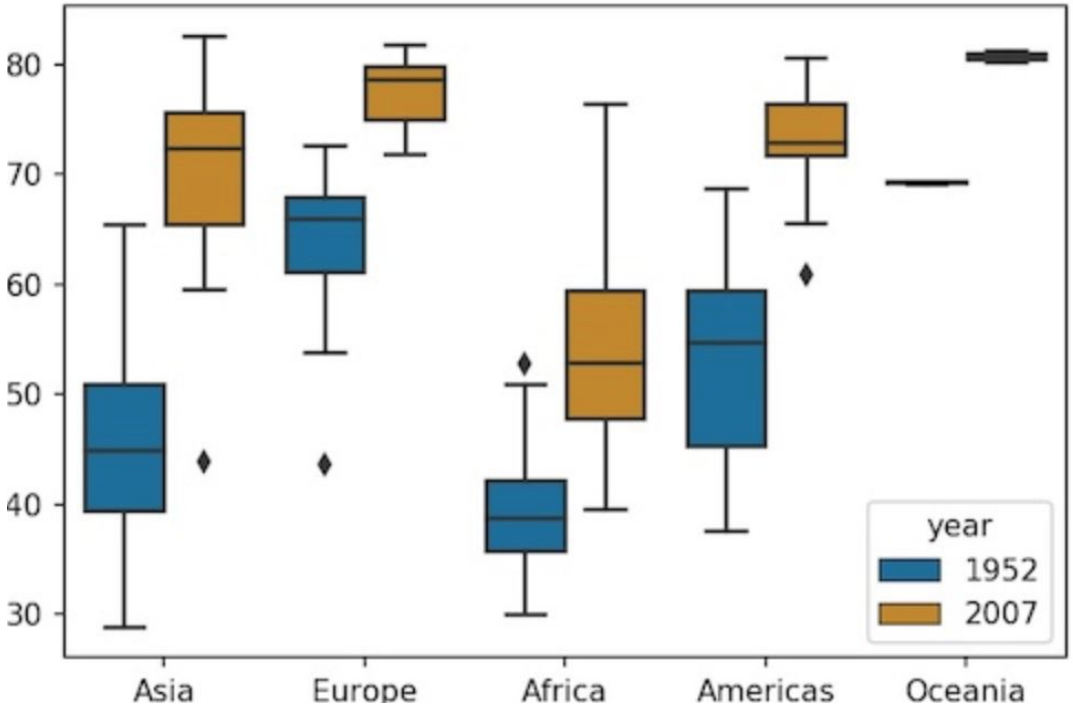
If we get a no...



SMALL REVIEW

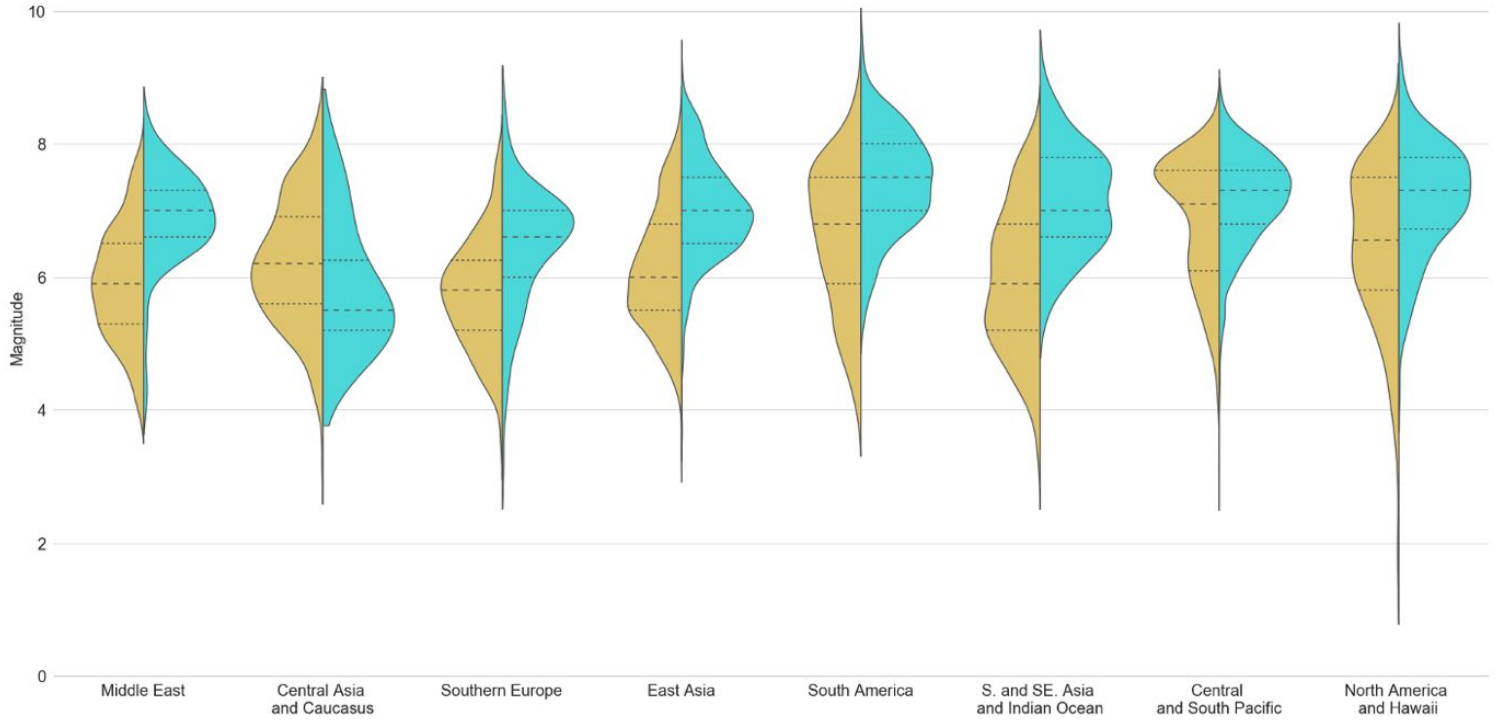
Saturdays.AI
Donostia

SMALL REVIEW: Boxplot

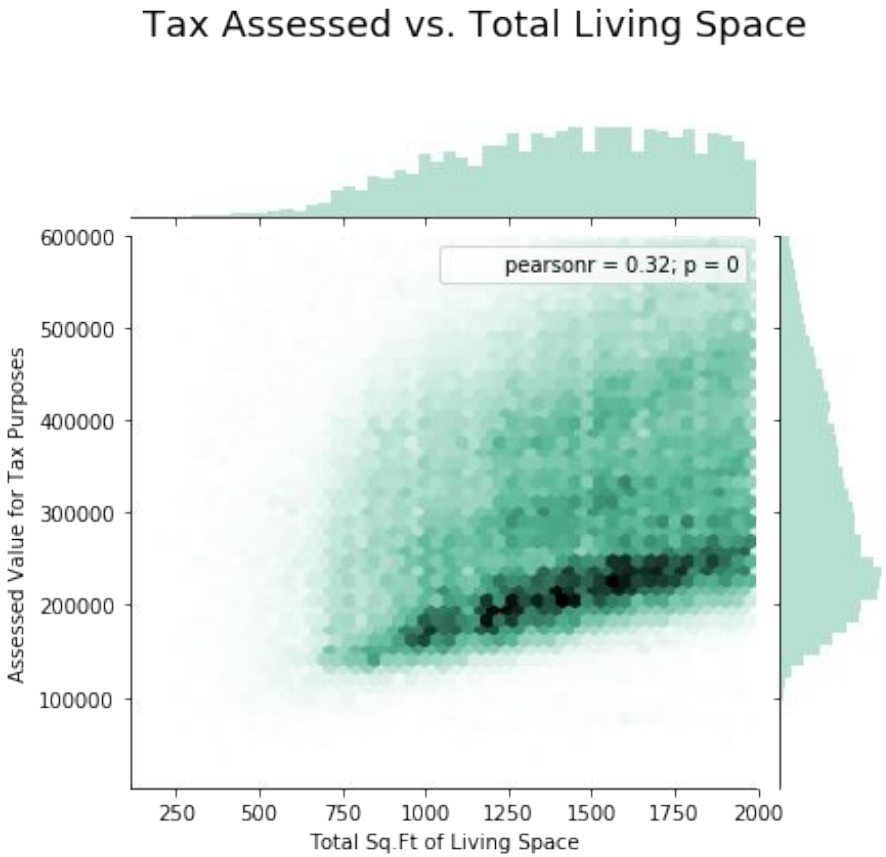


SMALL REVIEW: Violin Plot

Saturdays.AI
Donostia



SMALL REVIEW: Hexagonal Binning



TIPS

Saturdays.AI
Donostia

TIPS

1. Scale
2. Color Code
3. Axis
4. Units
5. Title
6. 0 **MUST APPEAR**

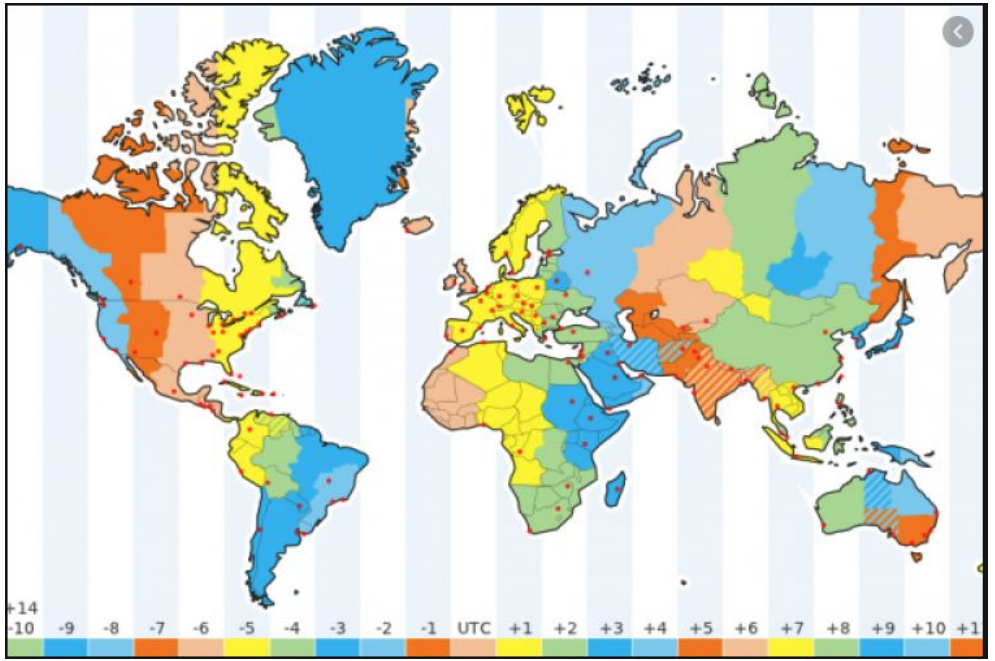
Saturdays.AI
Donostia

TIPS

Saturdays.AI
Donostia

Timestamp:

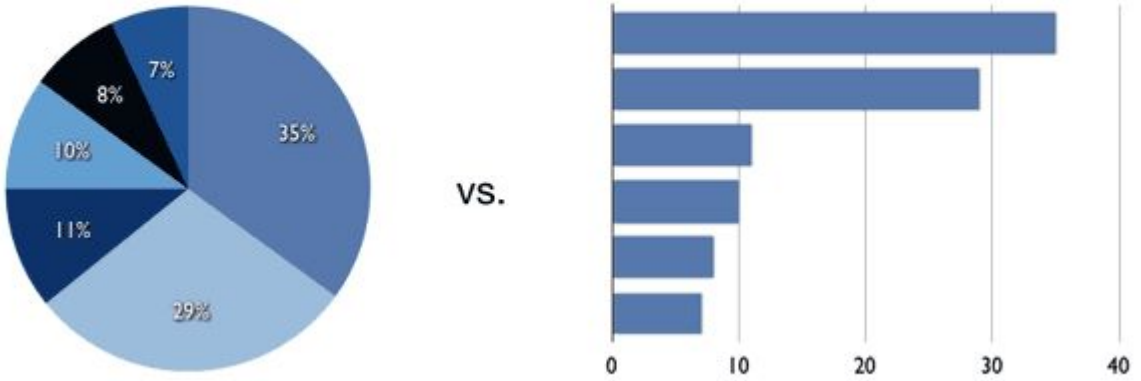
- Milliseconds since 01/01/1970
- Keep an eye on Timezones
  - Winter/Summer



TIPS

Saturdays.AI
Donostia

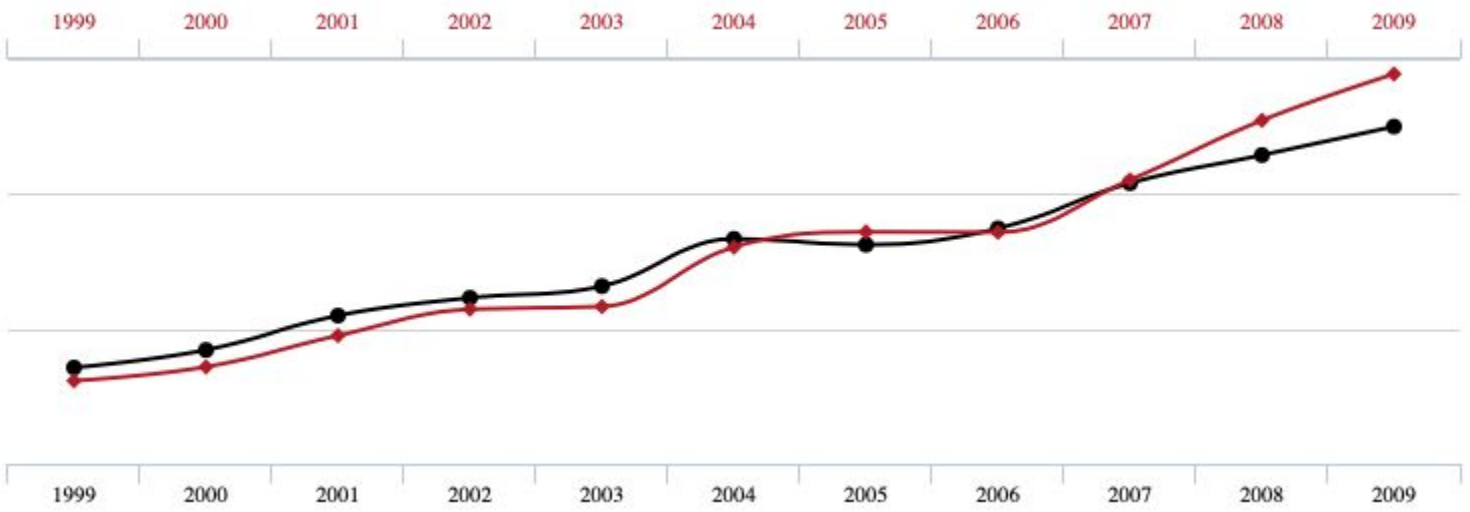
Pie Chart Vs. Bar Graph



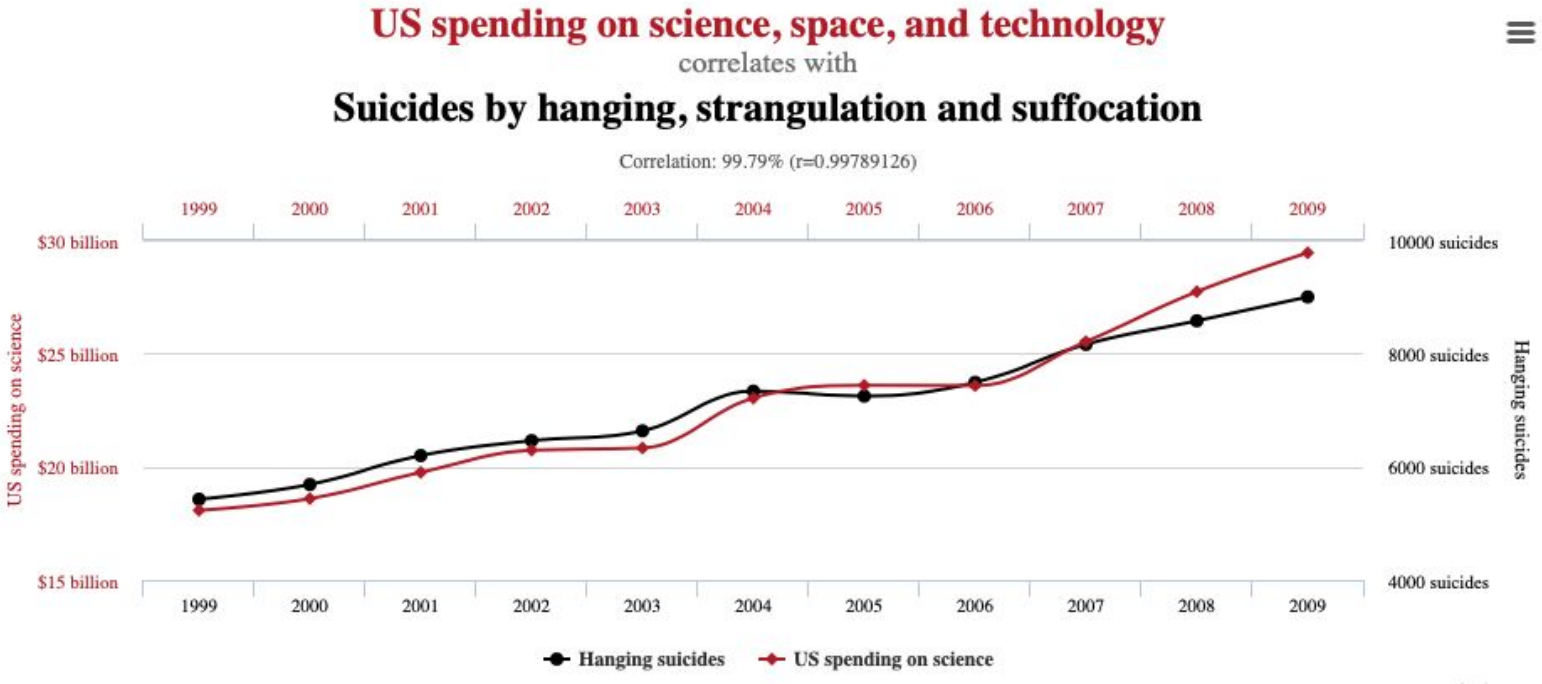
DO NOTS

Saturdays.AI
Donostia

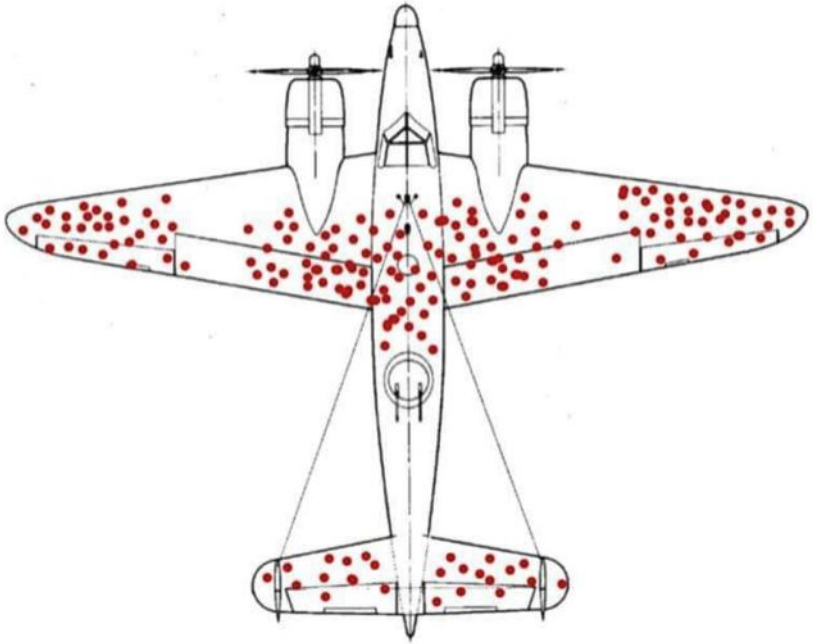
DO NOTS: Common mistakes I



DO NOTS: Common mistakes II



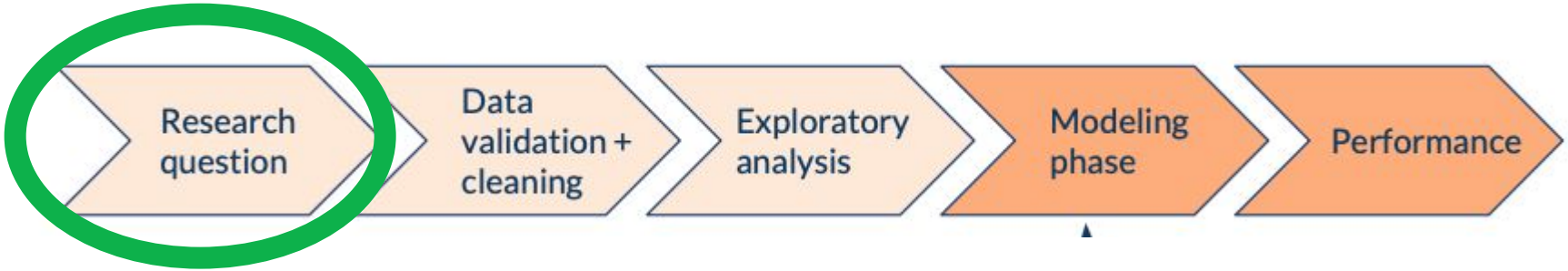
DO NOTS: Common mistakes III



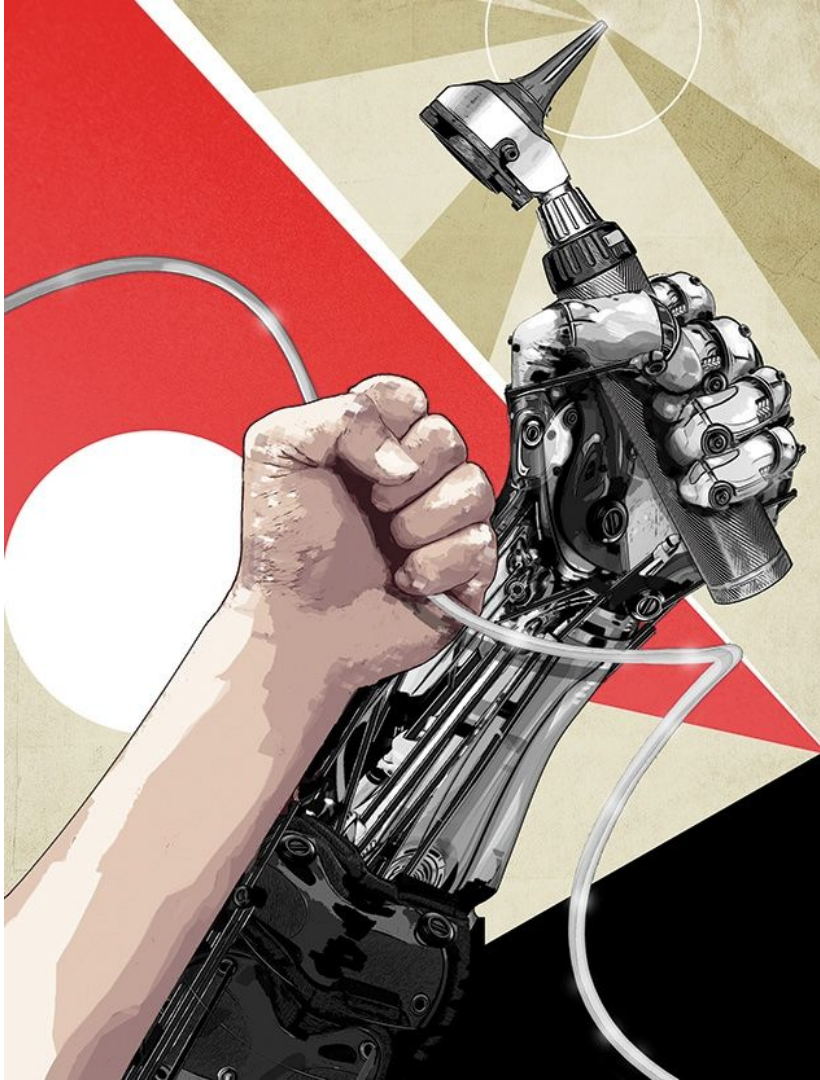
Saturdays.AI
Donostia

DO NOTS: Common mistakes IV

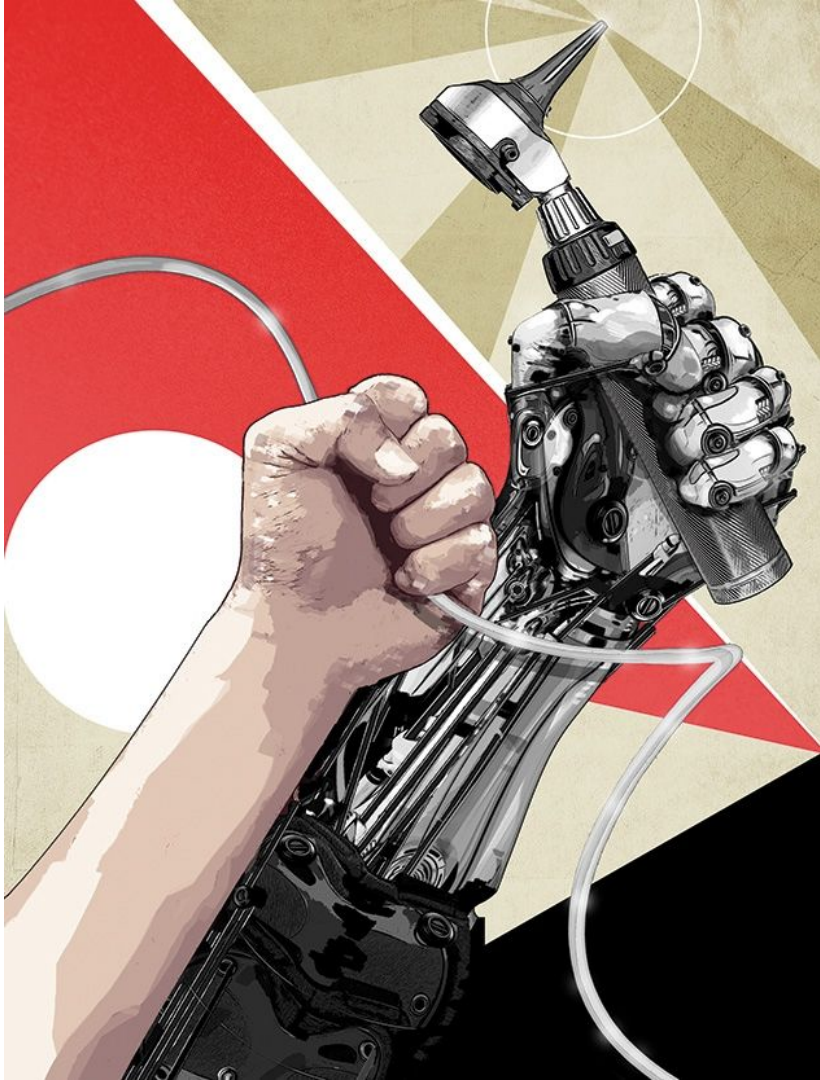
Saturdays.AI
Donostia



Practice - Time Series!



Challenge - Time Series!



Bibliografía

Saturdays.AI
Donostia

/1./ /Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow/

/2./ /Fast.AI - Introduction to Machine Learning for Coders/

/3./ /MLCourse.AI/

/4./ /DeltaAnalytics/

/5./ /The Hundred-page Machine Learning Book/

/6./ /Machine Learning for Humans (Vishal Maini)/

/7./ /Datacamp/

/8./ /DataQuest/

Partners

Saturdays.AI Donostia

Agradecemos a nuestros partners por confiar en nosotros para facilitar la formación en IA de cara a la 4ª Revolución Industrial.

IMPACT HUB

Bizkaia foru aldundia diputación foral

spri ENPRESA DIGITALA

eureka! zientzia museoa kutxa FUNDAZIOA

Deusto Universidad de Deusto

kutxa kultur kutxa FUNDAZIOA

eman ta zabal zazu UPV EHU

Celera Aceleradora de personas

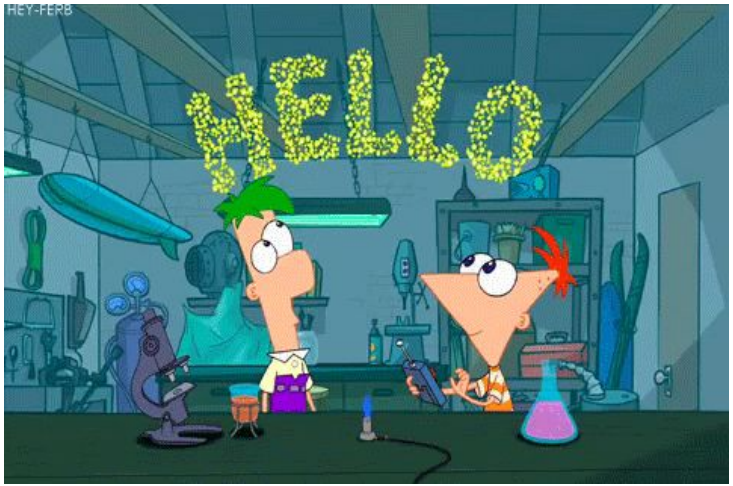
NVIDIA INCEPTION PROGRAM

aws

4YFN

Saturdays.AI
Donostia

WELCOME!



Saturdays.AI

This model fits me
95% of the time

CAMP TECNOLÓGICO

GLOBAL SHAPERS COMMUNITY BILBAO

www.saturdays.ai

donostia.saturdays.ai

donostia@saturdays.ai